# Analiza si vizualizarea datelor

Nicoleta ROGOVSCHI

nicoleta.rogovschi@parisdescartes.fr

# Self Organizing Map (SOM)

# Outline

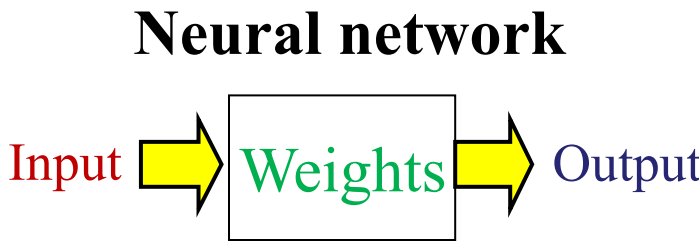
- Introduction and definitions of neural networks
- Introduction SOM
- Algorithm
- Example
- Conclusions

# Neural networks

- A set of interconnected neurons/information processing units
- A technique designed to model how the brain performs a particular task
- Used to extract the pattern of information from data sets where numbers are vast and has hidden relations
- Ability to handle noisy data

# Neural network learning

- By learning we can extract information
- This information is stored on the links between the neurons (also called weights)




- Using neural networks can perform two types of learning:

- Supervised
- Unsupervised

# Self-organisation

- The brain cells are self organizing themselves in groups, according to incoming information.
- This incoming information is not only received by a single neural cell, but also influences other cells in its neighbourhood.
- SOMs mechanism is also based on this principle

# The principle of SOMs

- SOM produces the similarity graph of the input data
- Converts non-linear relationships between high dimensional data into simple geometric relationships

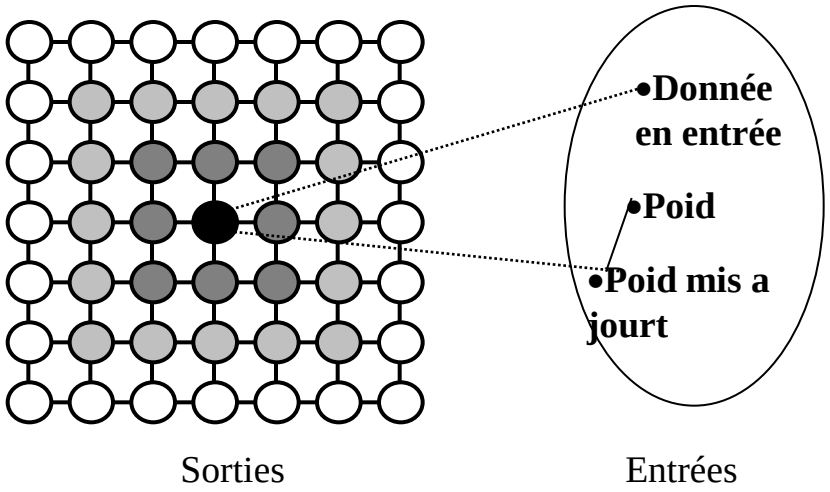



**Illustration d'une carte SOM de taille 7x7**

# SOM

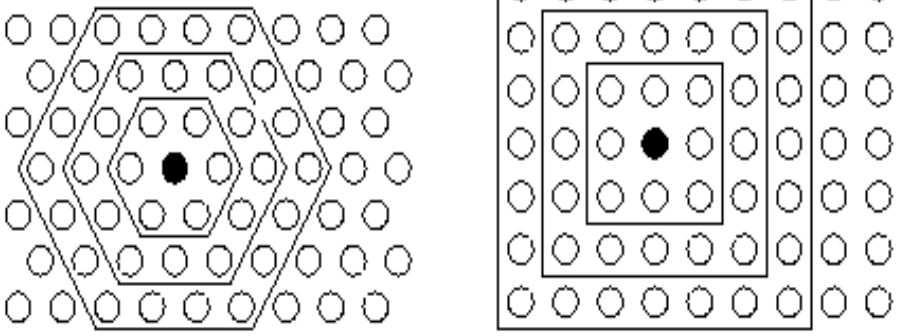
- Introduced in 1984 by Teuvo Kohonen
- Vector quantization + vector projection
- Technique based on unsupervised learning
- Used in classification and visualization of large data sets
- Used in many areas

# Fields of application

- Applications
  - Exploratory data analysis, clustering
  - Quantification, variable selection, outlier detection
  - Diagnosis, prediction, missing data
- Domains
  - Socio-economic
  - TextMining
  - Telecomunication
  - Industrial Processes

# SOM Architecture

- Set of neurons / cluster units
- Each neuron is assigned with a prototype vector that is taken from the input data set
- The neurons of the map can be arranged either on a rectangular or a hexagonal lattice

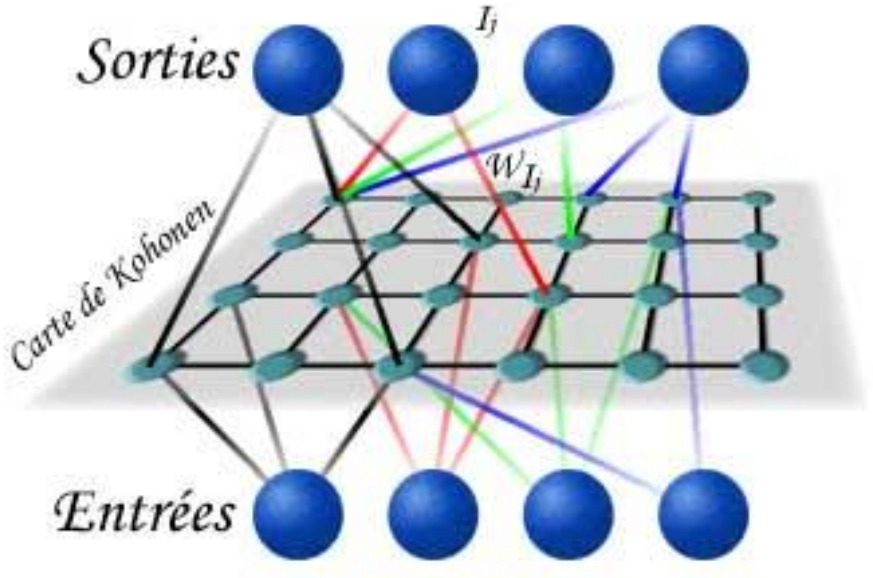


Hexagonal neighborhood  Rectangular neighborhood

# SOM

- Cost function:

$$C(w) = \sum_i \sum_k h_{k,s(x_i)} \left\| x_i - w_k \right\|^2$$



# SOMs Algorithm

1. Initialisation step
   - Define the topology map
   - Randomly initialize all the prototypes for each neuron
2. Competition step
   - Have a given datum $x_i$ randomly chosen
   - Determine the winning neuron according to the rule:

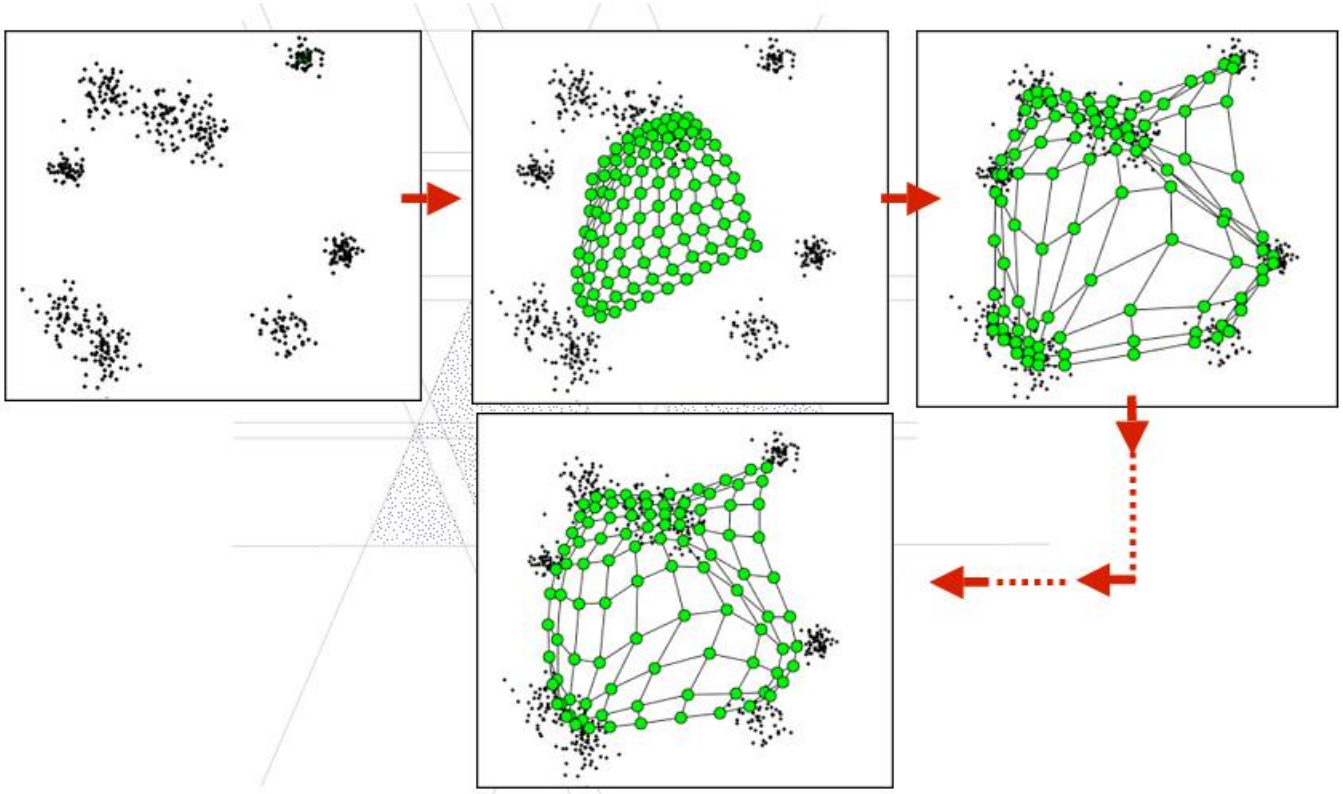
$$s(x_i) = Arg \min_{1 \le k \le m} \left\| x_i - w_k \right\|^2$$

3. Adaptation step
   - Adapt prototypes under rule

$$w_k(t+1) = w_k(t) + \varepsilon(t) h_{k,s(x_i)}(x_i - w_k(t))$$

4. Repeat steps 2 and 3 until the updates prototypes are insignificant

# Self organizing step



# Self organizing step

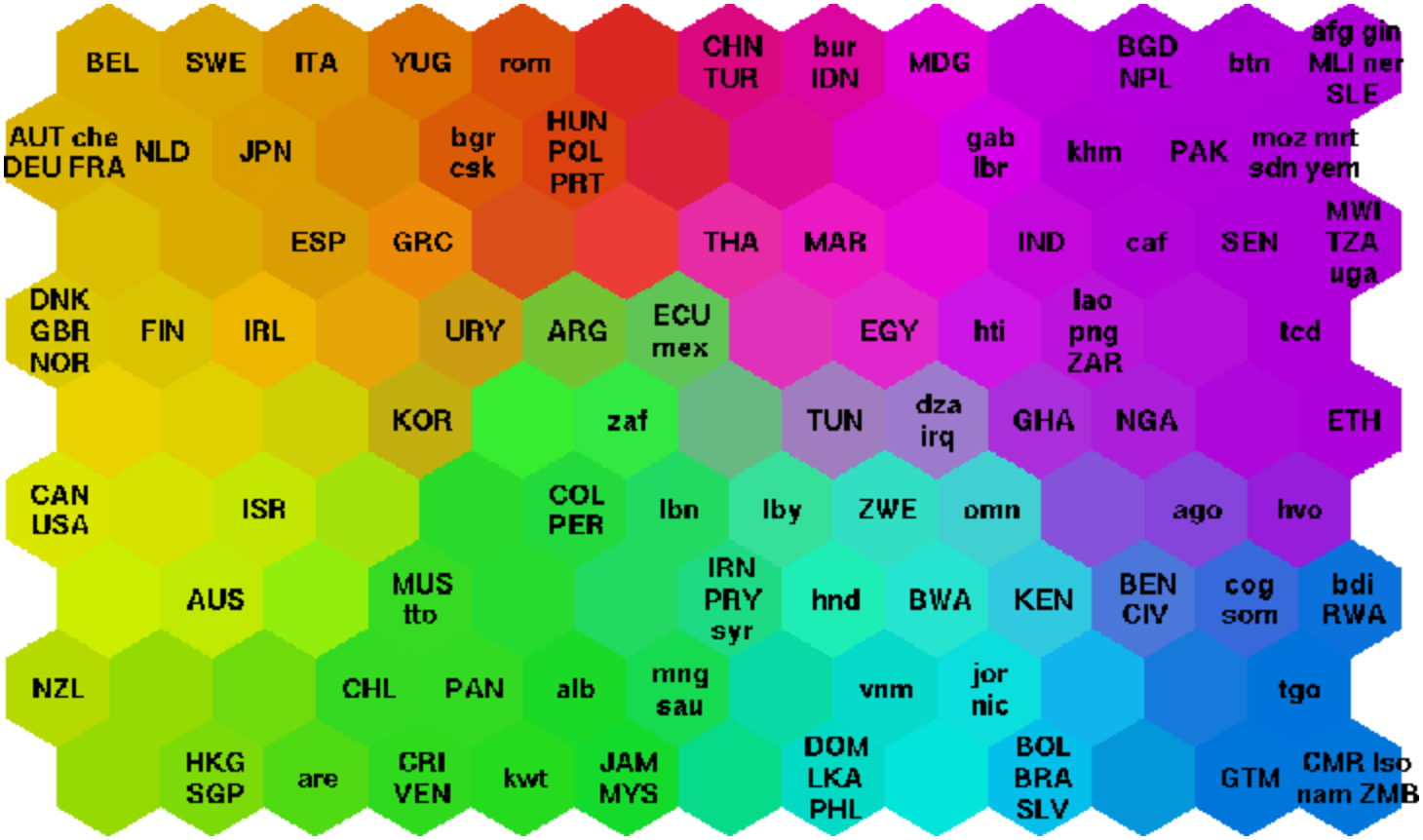
The map is organized by checking the following properties:

- Each neuron specializes in a portion of the input space
- Similar data have near projections on the map

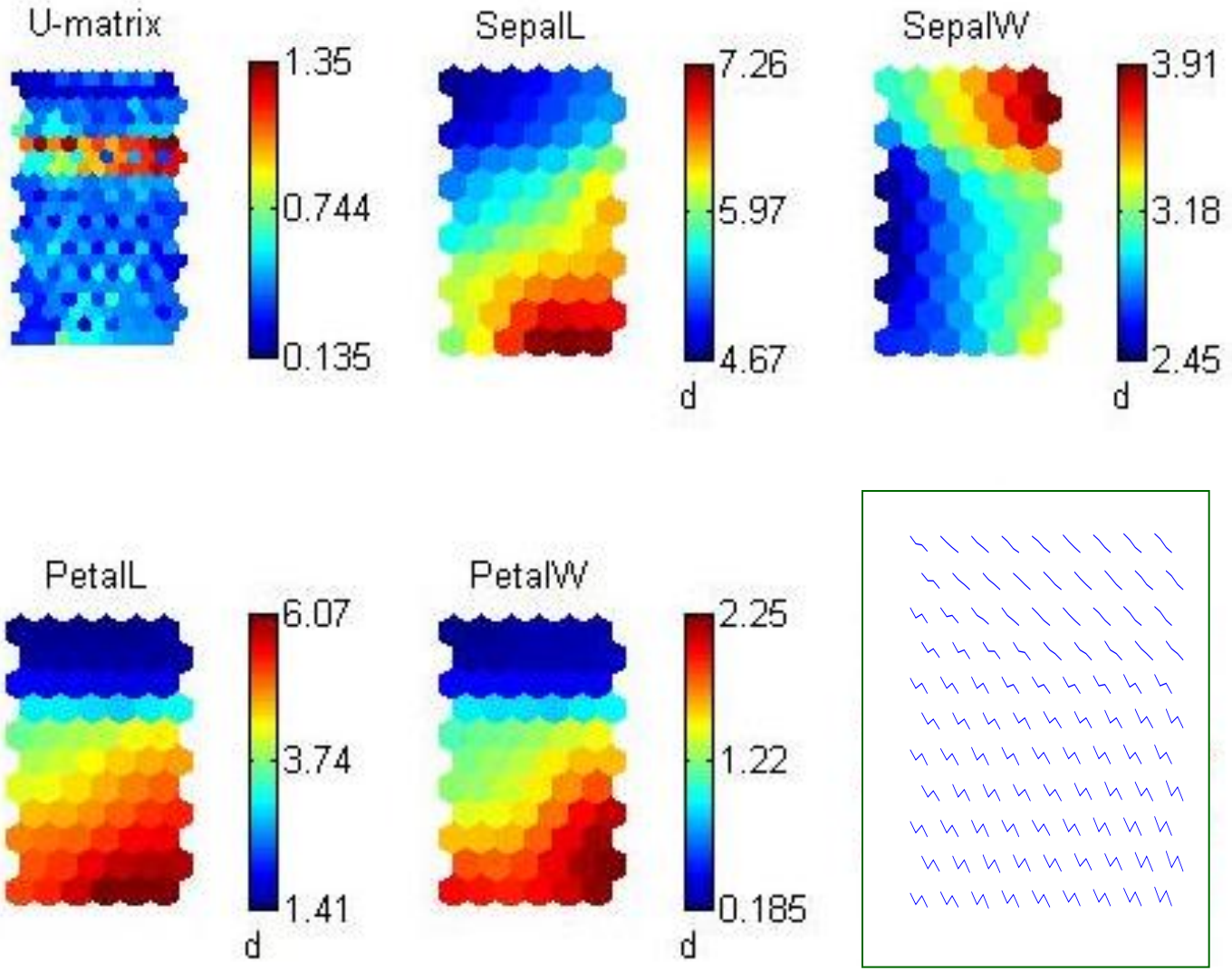
# Data visualization using SOMs

- Visualization of clusters and shape of the data (projections, U-matrices and other distance matrices)
- Visualization of components/variables (scatter plots, components planes)
- Visualization of data projections

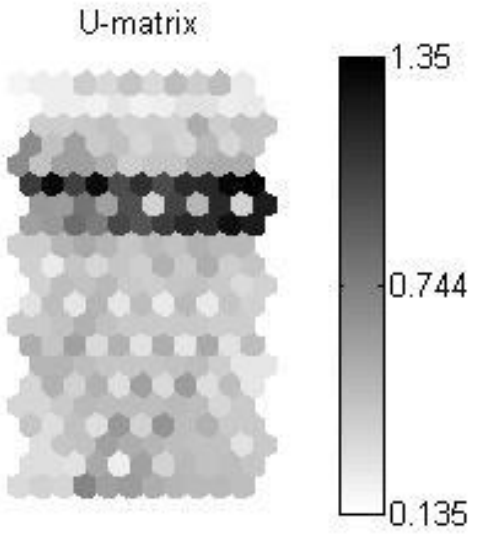
# Example : Countries of the world



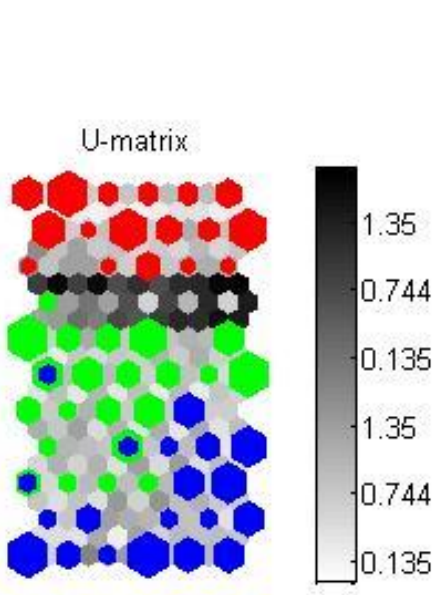
# Example: Iris



# Example: Iris




# Example: Iris




# Example: Iris